Jun 5th, 12:00 AM

# A method for managing the information resources in an interdisciplinary field

Sara von Ungern-Sternberg
*Äbo Akademi Universit*

# A METHOD FOR MANAGING THE INFORMATION RESOURCES IN AN INTERDISCIPLINARY FIELD

*Ungern-Sternberg, Sara von (formerly Sara von Flittner)*
Åbo Akademi University, Finland

---

The study reported here is based on a project addressing issues of collection development in interdisciplinary fields[1]. The aim of the project was to show that it is possible to use a small collection of articles to get a picture of the research specializations and of the subject structure in a field. The tool developed in the project enables librarians to identify a representative set of journals, and provides a tool for investigating the use of literature in the subfields.

Collection development in interdisciplinary fields is a challenge to librarians who are planning information provision. New fields emerge and scientists use literature from many disciplines and build up their own cores of literature gradually. Still the development of any subject field depends on how well it can make use of its knowledge base. To identify and organise information sources, for instance the production of literature, are therefore important issues in the development of new subject fields. How can the librarian decide which material should be included in the collections when even the scientists of the field have little concensus about the definition of the field? How can the librarian build up collections in fields where the classification systems, used for instance by journal services, have a discipline-based structure and include new interdisciplinary subject fields only at a very basic level, if at all. What methods are there to define a documentation language which is up-to-date and reflects the content of local research? These were the research questions for the project. The goals were to develop a method (a tool) that would help librarians plan the information provision in interdisciplinary fields and to find a method for selecting terms for a documentation language in such fields.

The tool was built up by the following methods:

1. The corpus of articles was analysed with regard to authorship, organizational affiliation and literature use.
2. The citations were analysed to get a picture of the journals. The subject fields of the cited journals were analysed according to their classification in Ulrich's.
3. The relationships between the cited journals was examined through a co-citation analysis.
4. The content of the articles was analysed by the indexing done for the different bibliographic databases. On the basis of this analysis a new documentation language that reflects the content of the research was constructed.
5. The knowledge structure of the field was examined using two methods; co-word analysis and bibliographic coupling.

**Material**

The interdisciplinary field chosen for this study was biotechnology and the study was limited to Finnish research in the field. Biotechnology is based on scientific research in biology, microbiology, biochemistry, biophysics, genetics, cellular biology, molecular biology, process engineering, industrial economy and other neighbouring fields. In the literature many various definitions on biotechnology can be found. Most of them agree that certain processes utilizing biological organisms are biotechnology, but it is not always as clear which processes do not belong to biotechnology. There is still confusion about the terminology of the field. Biotechnology literature is growing fast, both in printed and in electronic form. There is a great number of new publications in the field and research results are published in established journals as well as in new ones. Besides the core, which is directly associated with biotechnology, relevant basic and applied research is found in the literature of many other disciplines. The most important classification systems, such as UDC and Dewey, have not regarded biotechnology as a specific field, but scattered the subject to different fields

in science, medicine, and technology. This makes it difficult to identify the relevant publications in the field.

The material used in the research project was a Finnish collection of 176 articles in biotechnology published 1987-1991. The articles were selected from four different international bibliographic databases (*Derwent Biotechnology Abstracts, Current Biotechnology Abstracts, Pascal Biotechnologie and Life Sciences Collection, Biotechnology*), which claim to cover all aspects of biotechnology. An article had to be included in at least two databases to be selected. The overlap between the four databases meant that of a total of 909 references with Finnish corporate address during the period, 176 references were common to at least two databases. These 176 references were the basis on which the tool was tested.

**Bibliometric methods**

Bibliometric methods make it possible to describe the content, structure and development of research and can provide a tool for identifying the core for a collection in new fields. Citations are one of the fundamental ingredients in bibliometric analyses, and they are the basis for *co-citation analysis and bibliographic coupling*. In co-citation analysis the data compiled are counts of the number of times two documents are jointly cited in later publications. The fact of having been cited together in the same new paper establishes a link between the earlier papers and the strength of the link is a function of the number of times that a pair of documents are cited together. Special kinds of co-citation analysis use authors or journals as units of analysis. Journals can thus be used as a unit for the study of the organisation of a subject literature through co-citation analysis. Co-citations of the published articles link the journals in which they were published. The journal title then represents the subject contents of all the articles it includes[2]. Bibliographic coupling is based on the assumption that two articles which both cite the same previously published article have something in common. Analysis of bibliographic coupling results in clusters of citing documents,

whereas co-citation analysis groups cited documents. In bibliometric terminology the citing articles create a *research front*; a cluster of cited documents is called an *intellectual base*[3]. *Co-word analysis* is based upon the analysis of the co-occurrence of the keywords used to index articles and other documents[4]. This method identifies so-called problematic networks[5] and is also useful for identifying the research front of a field.

## Results

*Bibliometric characteristics*

The bibliometric characteristics of biotechnology in Finland show that the eleven most productive authors account for almost 70% of the articles. Every article has on average three authors, which indicates a high degree of cooperation between researchers in the field. Almost one fifth of the articles is the result of cooperation with researchers in other countries. International cooperation is very strong in the fields of forest industry, waste disposal, and genetic engineering. The most important countries for cooperation are USA and Germany. The most important organisations are universities, which account for more than 40% of the articles. About 19% of the articles have the corporate address of a commercial company, which supports the conclusion that biotechnology is a field with strong research and development elements and that basic and applied research are closely interlinked. The most productive organisation in Finland is a governmental research institute. Literature use, as reflected in the articles by citations, indicates that the scientific journal is the most important source of information (84% of the citations were from journal articles) in biotechnology, as in other fields of science and technology. The number of citations is on the average high and resembles the citation behaviour in sciences more than that of the citation behaviour in technology. The cited journals have an obsolescence factor of 0,85. According to a study of *Science Citation Index* the obsolescence factors in the fields of science and technology varied between 0,80 and 0,89[6]. According to this the

obsolescence was faster in special fields than in broad, heterogeneous fields, which indicates that biotechnology is relatively heterogeneous with an average obsolescence.

**The journal use**

An analysis of the cited journals shows that only six journals are needed to cover one-fourth of the cited articles, but 21 journals are needed to cover half of the articles. To cover all 3337 citations to articles, 423 journals are needed. A comparison with other scientific and technical fields shows that the core of journals is very small and the scattering to different journals is also relatively small. The literature in common is well-defined and published in journals in microbiology, biochemistry and biophysics, biology, chemical engineering and genetics. The six most cited journals are classified in Ulrich's as genetics, biotechnology, bacteriology and biochemistry.

An analysis of the relations between journals on the basis of co-citations[7] shows that the most co-cited journals with more than 40 co-citations are *Gene, Proceedings of the National Academy of Sciences*, *Nature* and *Journal of Bacteriology*. The emphasis on genetic engineering and genetics in biotechnology can be seen in *Gene*, but *Proceedings of the National Academy of Sciences* and *Nature* are general journals which are usually found in any library collection in the sciences.

A graph of the distribution of the co-cited journals between different clusters shows that only highly co-cited journals give clearly analysable groups. With a threshold of at least 26 co-citations four clusters, totalling 12 journals, can be identified.

1. *Nature*
   *Gene*
   *Proceedings of the National Academy of Sciences*
2. *Journal of Bacteriology*
   *Applied Microbiology and Biotechnology*
   *Analytical Biochemistry*
3. *Applied and Environmental Biotechnology*
   *Journal of Biological Chemistry*
   *Molecular Cloning* (a monograph)
4. *Journal of Molecular Biology*
   *Methods in Enzymology*
   *Nucleic Acids Research*

The clustering analysis above shows only the core of co-cited journals and the relationships between them. To get a picture of the relationships between journals on the next level, that is the less productive journals, the core is removed *(Gene, Proceedings of the National Academy of Sciences, Nature, Journal of Bacteriology* and the monograph *Molecular Cloning*). This enables a study of the relationships between journals on the next level. 11 journals form three clusters when the five most co-cited journals are "cut" out and a threshold value of 18 co-citations is used. One cluster of five journals includes only journals in biochemistry on this level:

1. *Journal of Biological Chemistry*
   *Analytical Biochemistry*
   *Nucleic Acids Research*
   *European Journal of Biochemistry*
   *Methods in Enzymology*

The second cluster includes journals in applied biochemistry:

2. *Agricultural and Biological Chemistry Tokyo* (agricultural aspects of biochemistry)
   *Enzyme and Microbial Technology* (technical aspects of biochemistry).

There is also a cluster with three journals in biotechnology and one in molecular biology:

3. *Applied and Microbiological Biotechnology*
   *Applied and Environmental Microbiology*
   *Journal of Biotechnology*
   *Journal of Molecular Biology*

To sum up the co-citation analysis: the core journals of the field include journals with a general coverage, as well as journals covering biotechnology, genetics, bacteriology, and biochemistry. The journals on the second level, the less productive journals, have a strong emphasis on biochemistry.

**Terms for a documentation language**

The subject content of the articles was analysed looking at the indexing done for the databases. A new set of index terms was constructed for every article, on the basis of the indexing in the databases. If at least two databases described the content of an article with the same concept, terms describing this were selected according to specific rules for the new set of terms. The rules used were based on the ones developed by Sievert and Verbeck (1987)[8] for converting terms to concepts.:

1. The terms are different grammatical or syntactical forms of the same concept (e.g. *information storage and retrieval* and *information retrieval*),
2. The term is a subordinated term to a more general term and there is a word in common (e.g. *reference services* and *library services*),
3. The terms are synonyms.

The new index terms contain 427 different terms; 60% of them occur just once, and 22 terms are, according to an estimated discrimination index, much-used terms that offer only weak discrimination power for information retrieval. The most frequent terms include general terms such as "enzyme", "fungi", "bacteria" and "gene", but

also very specific bacteria used in genetic engineering and microbiology. Nine of the 22 most frequent terms describe methods or techniques. The frequencies of the terms indicate that the field is still developing and that there is a high degree of interim research. No term is highly frequent if the level for high frequency is set at half of the articles[9]. The stable part of the research in biotechnology in Finland includes the term "enzyme". The largest categories of terms are microorganisms, methods, enzymes, chemical substances and terms in genetic engineering.

The index terms selected were used as the basis for a co-word analysis, i.e. the analysis of the co-occurrence of index terms in different articles.

**The knowledge structure**

The research front of biotechnology was studied by two methods; co-word analysis and bibliographic coupling. Bibliographic coupling is based on the assumption that citations reflect subject content in the same way as index terms and it can be used for identifying the research front in the same way as co-word analysis. The two methods complement each other and give together a picture of the knowledge structure in Finnish biotechnology.

In the analysis different thresholds were defined by the use of the software Bibmap. Bibmap is a software developed at Umeå University for analysis of small collections of articles. The thresholds are either for the number of common index terms, or for cited documents in two articles, or for a "normalised" coupling strength. This takes into account the different variations in the number of index terms or citations per article and their frequency in the whole collection of articles.

In order to examine the differences and their significance for developing library collections, the clusters based on citations were compared with the clusters based on index terms at the same thresholds. About 90 articles were mapped into a little more than 20 clusters in both cases. The topics that the clusters contain constitute the

research front in Finnish biotechnology. The co-word analysis method, however, gives more specific topics than bibliographic coupling. According to Bricker[10] every discipline has a structure derived from many related subject fields with common characteristics with varying levels of hierarchical integration. Every discipline has elements of heterogeneity and homogeneity. An analysis of the strong elements can be identified with a Bradford analysis of the clustered articles in the research front. This analysis indicates that biotechnological research in Finland is heterogeneous with a great scattering of the clusters (research topics). The most important research topics are:

- plant cell cultures
- waste disposal
- production of acids
- production of enzymes, biocatalysis
- cyclodextrines
- genetic engineering
- bacteria in milk and food packages
- genetically modified yeasts

The journal use in the research fronts is, to a large extent, different from the journal use in the whole collection of articles. In the research topic "Plant cell cultures" journals in botanics are much used. Four of the five core journals are botanical and the fifth is in the field of cellular biology. This cluster of articles shows a high conformity and differs in its use of journals from the field as whole. Just one of nine journals is included in the 30 most cited journals of the original 176 articles. The research topic "Waste disposal" shows high usage of journals in bacteriology. Two of the four core journals belong to this field. The two remaining journals are also classified as microbiology.

## Applicability and generalisability of the method

The developed tool provides a tool for collection development in an interdisciplinary field. It includes, though, weaknesses which the librarian must be aware of.

The *coverage* of the databases is one problem. Every database producer selects the primary documents in different ways. There can be considerable overlap between databases in the same field, but also gaps in coverage which are difficult to identify for the user. These gaps may be found in a group of databases. Biotechnological medicine is not covered in the articles, although it is a strong area in Finnish biotechnology.

The *citing practice* is another critical factor. Citations are a fundamental part of the tool and it has the deficiencies inherent in citing practice. Nevertheless, citations are still a fundamental instrument for judging the use of literature in different subject fields.

There were many problems in building up the terminology for a documentation language. Different databases use different depth and specificity for their *indexing*. The way in which special issues of a journal devoted to a particular theme are handled varies; they may be considered as one collective reference or many individual articles.

The two methods for *mapping the knowledge structure* have certain weaknesses. The results of the two methods differ somewhat. This can partly be explained by the roughness of both methods. The indexing language is built as a hierarchical structure which is not recognized in the co-word analysis. The bibliographic coupling is based on cited journals and is based on the assumption that the articles are aggregated to journals in specific fields. A scientific journal can, however, cover many fields and the content of it may change with time. The methods used are quantitative and neglect to some extent qualitative aspects of research. As a result the methods show the

knowledge structure based on the number of articles on a topic, but tell us little about the depth or value of the research.

**Discussion**

The heterogeneity which has been reflected in the study can be explained if biotechnology is considered as a collection of methods rather than a special discipline. The scientists are connected through common methods which they apply in other fields of science, medicine, agriculture or technology. Biotechnology is not an interdisciplinary field in the sense defined by Tengström[11]: "A knowledge process where new knowledge is sought from a number of disciplines". Biotechnology still has characteristics typical of a discipline. The field has concepts, theories, a corpus of literature and an organisation with research institutes and university departments. This type of interdisciplinarity finds support in literature. Tijssen[12] defines the concept interdisciplinarity as direct or indirect use of knowledge, *methods and techniques*, which have resulted from scientific or technological activities in other fields.

The knowledge structure of biotechnology is built up by research topics with their knowledge bases in other subject fields. These topics are often related through common scientists or research organisations. The project has shown that the use of journals in different research topics varies greatly and that scientists in biotechnology often go back to the knowledge base of their original field and use journals in fields other than biotechnology. The use of journals by local researchers thus determines how a collection should be developed.

The scattering of cited articles between different journals shows that the core of the journals is small, but the general distribution of articles between different journals is also small. This indicates that the common literature is well-defined and published in specific journals. The results of the project can be used for collection development in two ways: Firstly on the macrolevel, where the bibliometric characteristics of the field

can be used to support decisions on types of literature in a library collection, age, subjects covered, and the number of journals needed to cover a certain part of the relevant articles in the subject. Secondly on the microlevel, i.e. for decisions on journal titles, researchers, organisations and individual, much-used documents.

The tool developed gives a stable basis for information provision and sound knowledge which is objective. However, the development of a collection has also to be adapted to local research traditions and the characteristics of the field at the microlevel.

**References**


1. von Ungern-Sternberg, Sara. *Verktyg för planering av tvärvetenskaplig informationsförsörjning*. En tillämpning på ämnesområdet bioteknik i Finland. [A tool for planning information provision in interdisciplinary fields, applied to the field biotechnology in Finland] Åbo: Åbo Akademis förlag, 1994. 318p. Diss.
2. McCain, Katherine W. Mapping economics through the journal literature: An experiment in journal cocitation analysis. *Journal of the American Society for Information Science* 42 (4) 1991: pp. 290-296.
3. Persson, Olle. The intellectual base and research front of JASIS 1986-1990. *Journal of the American Society for Information Science* 45 (1) 1994: pp. 31-38.

Callon, Michel, Courtial, Jean-Pierre, Turner, William A., Bauin, Serge. From translations to problematic networks: An introduction to co-word analysis. *Social Science Information* 22 (2) 1983: pp. 191-235.

5. Courtial, J.-P., Callon, M. Is indexing trustworthy? Classification of articles through co-word analysis. *Journal of Information Science* 9 1984: pp. 47-56.
6. Griffith, Belver C. et.al. The aging of scientific literature. *Journal of Documentation* 35 (3) 1979: pp. 179-196.
7. Persson, Olle. *Forskning i bibliometrisk belysning*. [Research in the light of bibliometrics]. Umeå:Inum 1991.

8. Sievert, Mary Ellen, Verbeck, Alison. The indexing of the literature of online searching: a comparison of ERIC and LISA. *Online Review* 11 (2) 1987: pp. 95-104.

9. Ajiferuke, Isola & Chu, Clara M. Quality of indexing in online databases. An alternative measure for a term discriminating index. *Information Processing & Management* 24 (5) 1988: pp. 599-601.

10. Bricker, Robert. Deriving disciplinary structures: some new methods, models, and an illustration with accounting. *Journal of the American Society for Information Science* 42 (1) 1991: pp. 27-35.

11. Tengström, Emin. Biblioteks- och informationsvetenskapen - ett fler- eller tvärvetenskapligt område? [Library and information science - a multi- or interdisciplinary field?] *Svensk biblioteksforskning* (1) 1993: pp. 9-19.

12. Tijssen, Robert W. *Cartography of science*: Scientometric mapping with multidimensional scaling methods. Leiden: DSWO Press, Leiden University, 1992. 94p.